

PAGE: 1

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/303,216DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

<p>This Raw Listing contains the General Information Section and up to first 5 pages.</p>

ENTERED

```

1  <110> APPLICANT: Kim, Josphe L
2      Morgenstern, Kurt A
3      Caron, Paul R
4      Lin, Chao
5      Vertex Pharmaceuticals Inc.
6  <120> TITLE OF INVENTION: CRYSTALS OF HEPATITIS C VIRUS HELICASE OR FRAGMENTS THERE
7      COMPRISING A HELICASE BINDING POCKET
8  <130> FILE REFERENCE: Sequence listing for VPI/97-101 CIP CON
9  <140> CURRENT APPLICATION NUMBER: US/09/303,216
10 <141> CURRENT FILING DATE: 1999-04-30
11 <150> EARLIER APPLICATION NUMBER: PCT/US98/16879
12 <151> EARLIER FILING DATE: 1998-08-13
13 <160> NUMBER OF SEQ ID NOS: 4
14 <170> SOFTWARE: PatentIn Ver. 2.0
15 <210> SEQ ID NO 1
16 <211> LENGTH: 1932
17 <212> TYPE: DNA
18 <213> ORGANISM: Hepatitis C virus
19 <220> FEATURE:
20 <221> NAME/KEY: CDS
21 <222> LOCATION: (4)..(1896)
22 <223> OTHER INFORMATION: Full length HCV NS3 coding sequence
23 <220> FEATURE:
24 <221> NAME/KEY: misc_feature
25 <222> LOCATION: (504)..(1896)
26 <223> OTHER INFORMATION: Helicase domain
27 <400> SEQUENCE: 1
28      atg gcg ccc atc acg gcg tac gcc cag cag acg aga ggc aag ctt ggg      48
29      Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Lys Leu Gly
30      1          5          10          15
31      tgt ata atc acc agc ctg act ggc cgg gac aaa aac caa gtg gag ggt      96
32      Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
33      20          25          30
34      gag gtc cag atc gtg tca act gct acc caa acc ttc ctg gca acg tgc      144
35      Glu Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys
36      35          40          45
37      atc aat ggg gta tgc tgg act gtc tac cac ggg gcc gga acg agg acc      192
38      Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr
39      50          55          60
40      atc gca tca ccc aag ggt cct gtc atc cag atg tat acc aat gtg gac      240
41      Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp
42      65          70          75
43      caa gac ctt gtg ggc tgg ccc gct cct caa ggt tcc cgc tca ttg aca      288
44      Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr

```

PAGE: 2

RAW SEQUENCE LISTING PATENT APPLICATION US/09/303,216

DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

45	80	85	90	95	
46	ccc tgc acc tgc ggc tcc tcg gac ctt tac ctg gtt acg agg cac gcc				336
47	Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala				
48		100	105	110	
49	gac gtc atc ccg gtt cgc cgt cgc ggt gat agc cgt ggt agc ctg ctg				384
50	Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu				
51		115	120	125	
52	tct ccg cgt ccg att tcc tac ctg aaa ggc tcc tcg ggg ggt ccg ctg				432
53	Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu				
54		130	135	140	
55	ttg tgc ccc gcg gga cac gcc gtg ggc cta ttc agg gcc gcg gtg tgc				480
56	Leu Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys				
57		145	150	155	
58	acc cgt gga gtg gcc aag gcg gtg gac ttt atc cct gtg gag aac ctg				528
59	Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu				
60		160	165	170	175
61	gag acc acc atg cgt tcc ccg gtg ttc acg gac aac tcc tct cca cca				576
62	Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro				
63		180	185	190	
64	gct gtt ccc cag agc ttc cag gtg gcc cac ctg cat gct ccc acc ggc				624
65	Ala Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly				
66		195	200	205	
67	agt ggt aag agc acc aag gtc ccg gct gcg tac gca gcc cag ggc tac				672
68	Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr				
69		210	215	220	
70	aag gtg ttg gtg ctc aac ccc tct gtt gct gca acg ctg ggc ttt ggt				720
71	Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly				
72		225	230	235	
73	gct tac atg tcc aag gcc cat ggg gtc gat cct aat atc cgc acc ggt				768
74	Ala Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr Gly				
75		240	245	250	255
76	gtg cgt aca att acc act ggc agc ccc atc acg tac tcc acc tac ggc				816
77	Val Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly				
78		260	265	270	
79	aag ttc ctt gcc gac ggc ggg tgc tca ggt ggc gct tat gat atc atc				864
80	Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile				
81		275	280	285	
82	att tgt gac gag tgc cac tcc acg gat gcc aca tcc atc ttg ggc atc				912
83	Ile Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile				
84		290	295	300	
85	ggc act gtc ctt gac caa gca gag act gcg ggg gcg aga ttg gtt gtg				960
86	Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val				
87		305	310	315	
88	ctc gcc act gct acc cct ccg ggc tcc gtc acg gta ccg cat cct aac				1008
89	Leu Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn				
90		320	325	330	335
91	atc gag gag gtt gct ctg tcc acc acc gga gag atc cct ttc tac ggc				1056
92	Ile Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly				
93		340	345	350	
94	aag gct atc ccc ctc gag gtg atc aag ggc ggc cgt cat ctc atc ttc				1104

PAGE: 3

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/303,216

DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

95	Lys Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe	
96		355 360 365
97	tgt cac tca aag aag aag tgc gac gag ctc gcc gcg aag ctg gtc gca	1152
98	Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala	
99		370 375 380
100	ttg ggc atc aat gcc gtg gcc tac tac cgc gga ctt gac gtg tct gtc	1200
101	Leu Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val	
102		385 390 395
103	atc ccg acc agc ggc gat gtt gtc gtc gtg gcg acc gat gct ctc atg	1248
104	Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met	
105		400 405 410 415
106	act ggc ttt acc ggc gac ttc gac tct gtg ata gac tgc aac acg tgt	1296
107	Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys	
108		420 425 430
109	gtc act cag aca gtc gat ttc agc ctt gac cct acc ttt acc att gag	1344
110	Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu	
111		435 440 445
112	aca acc acg ctc ccc cag gat gct gtc tcc agg act cag cgc cgt ggt	1392
113	Thr Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly	
114		450 455 460
115	cgt acc ggc cgt ggg aag cca ggc atc tac aga ttt gtg gca ccg ggg	1440
116	Arg Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly	
117		465 470 475
118	gag cgc ccc tcc ggc atg ttc gac tcg tcc gtc ctc tgt gag tgc tat	1488
119	Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr	
120		480 485 490 495
121	gac gcg ggc tgt gct tgg tat gag ctc acg ccg gcg gag act aca gtt	1536
122	Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val	
123		500 505 510
124	cgt ctg cgc gcg tac atg aac acc ccg ggg ctt ccc gtg tgc cag gac	1584
125	Arg Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp	
126		515 520 525
127	cat ctt gaa ttt tgg gag ggc gtc ttt acg ggc ctc acc cat atc gat	1632
128	His Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp	
129		530 535 540
130	gcc cac ttt ctg tcc cag aca aag cag agt ggg gag aac ttt cct tac	1680
131	Ala His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro Tyr	
132		545 550 555
133	ctg gta gcg tac caa gcc acc gtg tgc gct cgt gcg caa gcc cct ccg	1728
134	Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro	
135		560 565 570 575
136	cca tcg tgg gac cag atg tgg aag tgt ttg atc cgc ctt aaa ccc acc	1776
137	Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr	
138		580 585 590
139	ctc cat ggg cca aca ccg ctc ctg tac cgt ctg ggc gct gtt cag aat	1824
140	Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn	
141		595 600 605
142	gaa gtc acc ctg acg cac cca atc acc aaa tac atc atg aca tgc atg	1872
143	Glu Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys Met	
144		610 615 620

PAGE: 4

RAW SEQUENCE LISTING PATENT APPLICATION US/09/303,216

DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

```

145      tcg gcc gac ctg gag gtc gtc acg ggatctggct cgcacatcatca tcatcatcac 1926
146      Ser Ala Asp Leu Glu Val Val Thr
147      625                      630
148      taatag 1932
149      <210> SEQ ID NO 2
150      <211> LENGTH: 631
151      <212> TYPE: PRT
152      <213> ORGANISM: Hepatitis C virus
153      <400> SEQUENCE: 2
154      Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Lys Leu Gly Cys
155      1 5 10 15
156      Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu
157      20 25 30
158      Val Gln Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile
159      35 40 45
160      Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr Ile
161      50 55 60
162      Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp Gln
163      65 70 75 80
164      Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr Pro
165      85 90 95
166      Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
167      100 105 110
168      Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
169      115 120 125
170      Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
171      130 135 140
172      Cys Pro Ala Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys Thr
173      145 150 155 160
174      Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu
175      165 170 175
176      Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
177      180 185 190
178      Val Pro Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
179      195 200 205
180      Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
181      210 215 220
182      Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
183      225 230 235 240
184      Tyr Met Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr Gly Val
185      245 250 255
186      Arg Thr Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys
187      260 265 270
188      Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
189      275 280 285
190      Cys Asp Glu Cys His Ser Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly
191      290 295 300
192      Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
193      305 310 315 320
194      Ala Thr Ala Thr Pro Pro Gly Ser Val Thr Val Pro His Pro Asn Ile

```

PAGE: 5

RAW SEQUENCE LISTING PATENT APPLICATION US/09/303,216

DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

```

195          325          330          335
196  Glu Glu Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys
197          340          345          350
198  Ala Ile Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys
199          355          360          365
200  His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu
201          370          375          380
202  Gly Ile Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
203  385          390          395          400
204  Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
205          405          410          415
206  Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
207          420          425          430
208  Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
209          435          440          445
210  Thr Thr Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg
211          450          455          460
212  Thr Gly Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu
213  465          470          475          480
214  Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
215          485          490          495
216  Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg
217          500          505          510
218  Leu Arg Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
219          515          520          525
220  Leu Glu Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala
221          530          535          540
222  His Phe Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro Tyr Leu
223  545          550          555          560
224  Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
225          565          570          575
226  Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
227          580          585          590
228  His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
229          595          600          605
230  Val Thr Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys Met Ser
231          610          615          620
232  Ala Asp Leu Glu Val Val Thr
233  625          630
234  <210> SEQ ID NO 3
235  <211> LENGTH: 8157
236  <212> TYPE: DNA
237  <213> ORGANISM: Artificial Sequence
238  <220> FEATURE:
239  <223> OTHER INFORMATION: Description of Artificial Sequence: Genetically
240  engineered plasmid containing full-length HCV NS3
241  coding sequence
242  <400> SEQUENCE: 3
243  ttaatacgac tcactatagg ggaattgtga gcggataaca attccccctct agaaataatt 60
244  ttgtttaact ttaagaagga gatataccat ggcgcccac acggcgtacg cccagcagac 120

```

PAGE: 6

VERIFICATION SUMMARY
PATENT APPLICATION US/09/303,216

DATE: 05/18/1999
TIME: 10:15:37

Input Set: I303216.RAW

Line ? Error/Warning

Original Text
